Running head: EFFECTS OF AYP REQUIREMENTS ON STUDENT ACHIEVEMENT

The Effects of AYP Requirements on Student Achievement

Canda D. Mueller
Applied Measurement Professionals

Vicki L. Schmitt
University of Kansas

Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA (April 2006).

Abstract

Three years of assessment data have been collected since The No Child Left Behind Act took effect. The purpose of this paper was to determine what effect adequate yearly progress (AYP) requirements have had on student achievement as measured by achievement level categorization across years. In measuring how well students are achieving in the wake of the AYP regulations, a method similar to the one suggested by Linn, Baker, and Betebenner in 2002 was used on Kansas assessment data from the years 1999-2000 through 2004-2005. Effect sizes were calculated using the proportions of students within an achievement level categorization across years for three grades assessed in reading and three assessed in mathematics.

The results showed mostly small effects for a time span of one year. Comparing data from one year prior to implementation of adequate yearly progress with 2004 and 2005, two and three years since implementation, effect sizes were typically medium or large. Regression toward the mean, indicated by not focusing on the instruction of high achieving students, did not seem to occur in this data set. In fact, the proportions of students in the two higher categories, *Advanced* and *Exemplary*, increased as proportions in *Proficient* remained relatively constant across the six years.

As is well documented and often discussed, the No Child Left Behind Act (NCLB) required states to develop adequate yearly progress (AYP) targets beginning with data from the 2002-2003 assessments (P.L. 107-110). While many authors feel that the law is unfair and enhances the achievement gap among racial/ethnic groups and between students that are economically advantaged and those that are not, the fact remains that all students are expected to reach proficiency as measured by state assessments by 2014. Three years of data have been collected since this law took effect. The question now becomes "Are we making adequate progress?"

Introduction

The Kansas Adequate Yearly Progress Guidance (KSBE, 2004) indicates that reading and mathematics state assessment results, assessment participation rates, and attendance and graduation rates are included in determining AYP. During the academic years 1999-2000 through 2004-2005, Kansas operated under consistent standards and assessments in grades 5, 8, and 11 for reading and 4, 7, and 10 for mathematics (KSBE, 1999a, 1999b). The AYP guidelines took effect in the midst of this assessment system such that the data from 1999-2000 through 2001-2002 was prior to the AYP targets and the data from 2002-2003 through 2004-2005 school years must meet AYP targets. The purpose of this paper is to determine what effect these new requirements have had on student achievement as measured by the proportion of students attaining the categorization of proficient or above as defined by the Kansas State Department of Education. The hypothesis is that while there have been gains in the proportions of students scoring in the proficient category, the overall increase will not show a large effect size in any grade level. Further, the proportions of students scoring in the two

higher categories, Advanced and Exemplary, on the state assessments will not vary a great deal from year to year.

Literature Review

Linn (2005), among others, have expressed concern about certain requirements that NCLB has for states to meet the AYP guidelines by 2014. Linn reports that many states have two accountability systems and that there is great confusion about what it means to meet the goals of one system and fail to meet the goals of the other. For example, he describes the state accountability system in Kentucky and what schools are also required to do to meet AYP guidelines. To address school accountability in the state system, assessments are given periodically to selected grades to reduce the testing burden. The results are then aggregated into a composite score to judge school accountability. The requirements put forth in NCLB maintain that all states will administer reading and mathematics assessments in grades 3-8 and once in high school every year and, further, compare growth on these assessments across cohorts from year to year. In 2004, 95.6% of Kentucky schools met the state goals, but only 74.0% met the AYP goal (Linn, 2005).

In contrast, other authors have remarked that NCLB's requirements have shown where gaps in achievement exist and have been successful in revealing problem areas in state education systems. Hall, Weiner, and Carey (2003) provide examples of schools across the nation that are performing as expected in helping all students succeed. Of course, this commentary came in the first year of implementation. As conceded by Kafer in 2004, little is yet known about how well the regulations are working. She goes on to note that there is indeed confusion among patrons regarding state and federal systems and

the meaning of progress in each one. Discrepancies in performance standards were recently highlighted in the summer 2005 policy brief published by the National Center for Research on Evaluation, Standards, and Student Testing (CRESST). CRESST specifically compared the differences in the percent of students in Colorado and Missouri scoring proficient and above on the 2003 National Assessment of Educational Progress (NAEP) as well as the percent of students scoring proficient and above as determined by each state's assessment program. While differences were noted in eighth grade math NAEP scores, with 34% of Colorado students scoring proficient or above compared to 28% of Missouri students, these differences are modest compared to the differences in the percent of students deemed by the state tests. In Colorado, 67% of students scored proficient or above on their state test but in Missouri only 21% of students were categorized as proficient above. Thus highlighting the vast differences that exist in the level of rigor associated with state performance standards (Linn, Summer 2005). As a result, the Missouri Board of Education recently changed their definition of proficiency from a term indicating student performance as above grade to student performance at grade level.

Another concern often voiced is that high achieving students are being left out in many instances because schools are so concerned about ensuring that all students meet the proficiency target by 2014 (Sanders, 2003). Sanders talks in particular about high achieving students in at-risk populations. In schools with high populations of at-risk students, high achievers tend to be overlooked because they tend to not be in jeopardy of missing proficiency goals (Sanders, 2003). He continues by arguing that all states should

go beyond AYP requirements and set growth standards for schools to meet with longitudinal data.

Adequate yearly progress measures for schools in most states use the change in the proportion of students meeting a proficiency target as a monitor for meeting the overall goal of 100% of students at proficient or above by 2014. These scores have been shown to be quite unstable at the school level (Linn & Haug, 2002). The authors in that article suggest combining data across years for more accurate information. It has been suggested that a reasonable alternative to this would be to use effect size as the metric for AYP purposes (Linn, Baker & Betebenner, 2002). Orlich (2003) did perform this operation on archival data from the Washington Assessment of Student Learning (WASL). He found no positive effect from year to year, but did find a small effect size when analyzing the data over a four year period (Orlich, 2003).

In order to measure how well Kansas students are achieving in the wake of the AYP regulations, a method similar to the one suggested by Linn, Baker, and Betebenner in 2002 and carried out by Orlich in 2003 was used on Kansas assessment data from the years 1999-2000 through 2004-2005. In this way, data was available for three years prior to the AYP regulation implementation and three years hence. Linn and his colleagues suggested the use of an effect size target as a way of moving away from performance level categorization so that smaller gains in achievement could be detected. In this study, the annual change in proportion was of interest. Effect sizes were calculated using the proportions of students within an achievement level categorization across two years and the standard deviation of this proportion away from the state average proportion for each grade and subject (5, 8, and 11 for reading and 4, 7, and 10 for mathematics). Kansas uses

five levels of achievement categorization, Basic, Unsatisfactory, Proficient, Advanced, and Exemplary (KSBE, 2004). The proportions of students in each achievement category within each school district were utilized in obtaining an effect size to analyze the change in proportional achievement across years. Data were used for each individual performance category so that regression toward the mean, as suggested by Sanders (2003) for example, was also studied.

Methodology

Over the course of six years, the number of school districts in the state of Kansas has changed because of school consolidation. As of 2005, there were 301 unified school districts in the state (KSDE website). However, not all districts contain enough students at a particular grade level to report achievement. Table 1 shows the number of districts used in the proportional calculations for each grade each year from 2000 through 2005. For example, 290 districts were used in the 11^{th} grade calculations in 2002, but only 285 districts had enough students for reporting purposes in 2003.

Table 1: Number of Students per Grade per Year

Year	4^{th}	5^{th}	7^{th}	8^{th}	10^{th}	11^{th}
2000	287	283	287	287	285	286
2001	285	287	281	287	290	284
2002	283	282	285	281	286	290
2003	280	283	288	286	286	285
2004	282	283	285	287	285	286
2005	283	285	286	286	286	283

Effect size was calculated by taking the difference in the proportions of students in each categorization from year to year and dividing by the standard deviation from the

state mean. This was done for consecutive years as well as across multiple years. The nominal levels of effect as defined by Cohen (1988) of at least 0.2 for a small effect, a minimum of 0.5 for a medium effect, and 0.8 or greater for a large effect were used for effect size comparison. The earlier year and accompanying standard deviation were used as the subtractend and denominator, respectively.

The mean proportions in each achievement level category and the accompanying standard deviations are shown in Table 2 through Table 7. As the years increase, the proportion of students scoring in the *Unsatisfactory* and *Basic* categories decreases in all grades, regardless of the subject. The proportion of students in the *Proficient* category stays relatively constant. Finally, the proportion of students in both *Advanced* and *Exemplary* increase from the results reported in 2000 to the results reported in 2005.

As these tables indicate, the mathematics assessments tend to be a bit more difficult for the examinees than the reading assessments at the middle and high school levels. For example, 44% of 7th grade students performed at the *Unsatisfactory* and *Basic* levels on the 2000 mathematics assessment. In 2005, this proportion had dropped to 28%. Approximately 31% of the eighth graders performed at the *Unsatisfactory* or *Basic* levels on the 2000 reading assessment. In 2005, about 21% of the eight graders were in one of these two categories. A similar phenomenon happened in tenth grade mathematics as compared to eleventh grade reading. In 2000, 58% of the tenth grade students were in the two lowest achievement levels, and this dropped to 45% in 2005. In eleventh grade 44% of students performed at the *Unsatisfactory* or *Basic* levels in 2000. This dropped to 34% by 2005.

Table 2: 4th Grade Mathematics Achievement Category Proportions and Standard Deviations

Year	U	U s.d.	B	B s.d.	P	P s.d.	A	A s.d.	E	E s.d.
2000	12.08	8.71	23.23	10.46	25.21	7.66	26.88	9.72	12.59	9.82
2001	9.13	7.13	20.07	10.38	26.24	8.78	27.69	10.81	16.87	12.37
2002	9.01	7.73	20.12	10.21	23.47	7.54	29.28	10.27	18.13	13.53
2003	6.38	6.14	16.82	9.94	22.42	9.80	30.72	9.57	23.64	15.32
2004	4.03	5.06	13.28	9.51	19.98	9.06	31.42	9.66	31.31	17.47
2005	2.93	3.96	9.48	7.82	16.86	9.15	33.87	9.35	36.84	17.64

Note. U=Unsatisfactory, B= Basic, P= Proficient, A=Advanced, E=Exemplary, s.d.= standard deviation

Table 3: 5th Grade Reading Achievement Category Proportions and Standard Deviations

Year	U	U s.d.	B	B s.d.	P	P s.d.	A	A s.d.	E	E s.d.
2000	12.01	8.17	24.13	8.02	22.99	6.81	25.95	8.71	14.91	7.53
2001	12.02	7.76	22.44	8.20	25.37	7.31	26.11	9.28	14.09	8.25
2002	11.72	7.95	25.56	8.92	24.18	7.17	24.59	9.27	14.07	8.01
2003	8.24	6.58	22.41	8.90	24.20	7.91	28.54	8.52	16.58	9.50
2004	7.10	6.72	21.08	10.08	23.20	7.62	28.65	8.94	19.80	11.53
2005	4.29	4.88	17.54	10.08	23.66	8.45	31.98	10.05	22.61	11.65

Note. U=Unsatisfactory, B= Basic, P= Proficient, A=Advanced, E=Exemplary, s.d.= standard deviation

Table 4: 7th Grade Mathematics Achievement Category Proportions and Standard Deviations

Year	U	U s.d.	B	B s.d.	P	P s.d.	A	A s.d.	E	E s.d.
2000	19.69	10.66	24.80	8.56	22.72	7.78	20.68	9.35	12.14	10.07
2001	19.02	10.96	22.86	8.40	22.87	6.94	20.94	8.47	14.32	11.10
2002	17.84	11.74	23.59	8.86	22.39	7.51	21.67	8.46	14.53	10.04
2003	15.88	10.09	21.13	8.51	23.30	8.19	23.50	9.07	16.18	10.32
2004	11.98	8.84	20.04	8.83	23.05	7.58	25.80	9.21	19.10	11.36
2005	9.18	7.86	18.66	9.52	22.63	8.18	27.02	8.74	22.51	12.74

Note. U=Unsatisfactory, B= Basic, P= Proficient, A=Advanced, E=Exemplary, s.d.= standard deviation

Table 5: 8th Grade Reading Achievement Category Proportions and Standard Deviations

Year	U	U s.d.	B	B s.d.	P	P s.d.	A	A s.d.	E	E s.d.
2000	9.70	7.26	21.56	8.14	29.85	7.67	30.81	9.61	8.42	5.95
2001	9.41	7.03	21.88	7.80	31.77	7.59	28.66	9.32	8.25	5.78
2002	10.41	7.16	22.12	8.13	30.13	7.80	29.27	9.87	8.05	5.45
2003	7.44	6.04	21.08	8.61	29.49	7.98	31.10	9.92	10.24	6.41
2004	5.44	4.95	18.95	9.04	29.82	7.73	34.69	9.77	11.13	7.14
2005	4.42	4.70	16.71	8.52	30.18	8.87	36.41	9.98	12.34	8.70

Note. U=Unsatisfactory, B= Basic, P= Proficient, A=Advanced, E=Exemplary, s.d.= standard deviation

Table 6: 10th Grade Mathematics Achievement Category Proportions and Standard Deviations

Year	U	U s.d.	B	B s.d.	P	P s.d.	A	A s.d.	E	E s.d.
2000	27.91	11.24	29.87	7.68	19.73	7.36	11.40	5.98	11.08	7.41
2001	24.88	11.85	29.87	9.23	21.31	7.83	11.45	6.61	12.42	9.03
2002	25.46	11.59	30.28	8.35	20.74	7.34	11.79	6.49	11.73	7.26
2003	24.17	11.02	28.98	8.80	21.29	7.60	12.46	6.36	13.18	8.09
2004	18.57	10.41	29.36	8.78	21.40	7.37	13.50	6.52	16.87	9.99
2005	20.22	12.40	24.84	8.55	21.81	7.47	14.72	7.56	18.40	10.67

Note. U=Unsatisfactory, B= Basic, P= Proficient, A=Advanced, E=Exemplary, s.d.= standard deviation

Table 7: 11th Grade Reading Achievement Category Proportions and Standard Deviations

Year	U	U s.d.	B	B s.d.	P	P s.d.	A	A s.d.	E	E s.d.
2000	15.20	8.91	28.94	8.77	25.39	7.73	21.05	8.36	9.41	5.61
2001	14.40	8.18	30.67	8.67	26.79	7.84	18.57	6.95	9.45	5.79
2002	15.88	9.03	29.57	8.28	25.19	7.85	20.48	7.66	8.92	5.44
2003	12.31	7.98	27.63	8.72	25.92	8.08	23.10	8.38	11.20	6.92
2004	10.36	7.08	26.61	8.69	25.74	8.03	25.42	8.38	11.86	6.51
2005	8.92	6.18	25.17	8.61	26.00	7.74	27.66	8.63	12.23	7.12

Note. U=Unsatisfactory, B= Basic, P= Proficient, A=Advanced, E=Exemplary, s.d.= standard deviation

Results

At least in Kansas, regression toward the mean does not seem to be evident. Students continue to move from the lower achievement level categorizations to the higher categorizations throughout the six years of testing. While there are many students categorized as *Proficient*, it is not because the proportions of students in *Advanced* and *Exemplary* decreased, but instead because the proportions in *Unsatisfactory* and *Basic* decreased. The purpose of this study was to determine if the change within each achievement category from year to year indicated a measurable effect size based on Cohen's definition (1988).

Seventy five effect size estimates were calculated for each grade. Approximately half of these estimates were classified as small or medium in each grade level. For comparison purposes, the number of effect size comparisons in each achievement categorization per grade is presented in Table 8. Note that effect sizes that are considered small or those that are negligible are most common for every grade. It is interesting to note that there are 19 effects that fit Cohen's definition of large in fourth grade, but most of those are between 3 years or more. Only one is large between two years, and that effect occurs between 2003 and 2005 for the proportion of students achieving at the Exemplary level. Every other effect that is considered large occurred between three or more years. Most were a year prior to AYP implementation compared to 2005, the last year for which data is currently available.

Table 8: Effect Size Estimates per Grade

Grade	None	Small	Medium	Large
4	9	30	17	19
5	26	26	14	9
7	30	24	18	3
8	29	24	18	4
10	38	20	14	3
11	29	28	15	3

Table 9 illustrates the effect sizes of students in each achievement level disaggregated by grade and subject for 2003 as compared to 2002. This is one year prior to AYP implementation compared to the year of AYP implementation. Note that effects are either small or do not exist in each grade and achievement level category with the exception of *Advanced* at fifth grade with an effect of -0.55. A negative effect, such as in the *Advanced* level for fifth grade indicates that the proportion of students in that category increased from 2002 to 2003. A positive effect, such as in the *Basic* level at tenth grade indicates that the proportion of students in that category decreased from 2002 to 2003. Therefore, it would be appropriate to expect positive effect size estimates in both the *Unsatisfactory* and *Basic* levels, while expecting negative effect size estimates in the *Proficient*, *Advanced*, and *Exemplary* categories.

Table 9: Effect Size Comparisons for 2002-2003

Grade	U e.s.	Effect	B e.s.	Effect	P e.s.	Effect	A e.s.	Effect	E e.s.	Effect
4	0.35	Small	0.34	Small	0.14	None	-0.15	None	-0.42	Small
5	0.44	Small	0.34	Small	0.07	None	-0.55	Medium	-0.36	Small
7	0.16	None	0.27	Small	-0.10	None	-0.19	None	-0.18	None
8	0.41	Small	0.10	None	0.08	None	-0.20	Small	-0.41	Small
10	0.11	None	0.17	None	-0.08	None	-0.11	None	-0.20	Small
11	0.41	Small	0.26	Small	-0.12	None	-0.34	Small	-0.44	Small

Note. U= Unsatisfactory, B= Basic, P=Proficient, A=Advanced, E=Exemplary, e.s.=effect size

Table 10 and Table 11 show the effect size estimates when the results for 2002 are compared to 2004 and 2005. That is, the results one year prior to implementation compared to one and two years after implementation. Most effects were medium based on Cohen's definitions (1988) when comparing 2002 results to 2004. An effect size is considered medium if the effect size estimate is between 0.5 and 0.8. The one effect deemed large occurred in the *Exemplary* category in the 4th grade mathematics results. Here, the proportion of students meeting this categorization almost doubled in this time span, from 18% in 2002 to 31% in 2004.

Table 10: Effect Size Comparisons for 2002-2004

Grade	U e.s.	Effect	B e.s.	Effect	P e.s.	Effect	A e.s.	Effect	E e.s.	Effect
4	0.65	Medium	0.67	Medium	0.46	Small	-0.20	Small	-0.98	Large
5	0.56	Medium	0.49	Small	0.13	None	-0.54	Medium	-0.71	Medium
7	0.52	Medium	0.41	Small	-0.08	None	-0.51	Medium	-0.47	Small
8	0.69	Medium	0.41	Small	0.03	None	-0.71	Medium	-0.55	Medium
10	0.58	Medium	0.14	None	-0.12	None	-0.25	Small	-0.69	Medium
11	0.60	Medium	0.36	Small	-0.08	None	-0.62	Medium	-0.53	Medium

Note. U= Unsatisfactory, B= Basic, P=Proficient, A=Advanced, E=Exemplary, e.s.=effect size

Table 11 shows the effect size estimates obtained when comparing 2002 results to 2005 results. This comparison shows many more effects that were deemed large, 0.8 or greater, based on Cohen's effect size estimate definitions (1988). As one would expect, the effect size for the *Exemplary* category for fourth grade mathematics was again large. In this comparison, the proportion changed from approximately 18% to 36% and did double in size. The proportion of students in the *Unsatisfactory* category in 5th grade reading decreased from 8% in 2002 to 4% in 2005. This was also a large effect.

Table 11: Effect Size Comparisons for 2002-2005

Grade	U e.s.	Effect	B e.s.	Effect	P e.s.	Effect	A e.s.	Effect	E e.s.	Effect
4	0.80	Large	0.67	Medium	0.46	Small	-0.42	Small	-1.41	Large
5	0.96	Large	0.89	Large	0.10	None	-1.05	Large	-1.08	Large
7	0.75	Medium	0.54	Medium	-0.04	None	-0.61	Medium	-0.80	Large
8	0.84	Large	0.66	Medium	0.02	None	-0.92	Large	-0.82	Large
10	0.47	Medium	0.64	Medium	-0.15	None	-0.45	Small	-0.93	Large
11	0.76	Medium	0.55	Medium	-0.12	None	-0.92	Large	-0.61	Medium

Note. U= Unsatisfactory, B= Basic, P=Proficient, A=Advanced, E=Exemplary, e.s.=effect size

Effect size estimates were also calculated for other year comparisons. Spans of three, four, five, and six years were calculated for research purposes. As one would suppose based on the results shown for 2002 in comparison to 2003, 2004, and 2005, larger effect size estimates occur between greater spans of time. Consequently, the greatest change is shown between 2000 and 2005 as anticipated.

Between any two years, one year of time, the effect was at most small. While schools continue to help more students achieve the categorization of *Proficient*, *Advanced*, or *Exemplary*, the proportional change between any two years is small. For the purposes of discussion for this presentation, the focus will remain on the time periods reported in this section.

Discussion

The number of students reaching the level of Proficient in the state of Kansas has indeed increased over the past six years. This study provides information on how significant the increase actually is. Between any two years, the increases in proportions of students meeting the requirements for *Proficient*, *Advanced*, or *Exemplary* showed at most a small effect. However, when comparing data from the year prior to federal implementation to 2004 data or 2005 data, effect size estimates were more likely to fall into the medium or large categories. This is the third year of data collection after AYP target implementation. While it is understood that comparing data from year to year is not necessarily best practice, it is the practice used by many states to meet the goal of 100% proficient by 2014. With three years of data, a trend is beginning to take shape. The research reported here shows where one state is and where it needs to go to be successful under the No Child Left Behind Act. That is, where it needs to go to ensure that 100% of students are proficient or above by 2014.

This study also analyzed the proportions of students in the categories above Proficient as defined by the state of Kansas. There is definitely concern among researchers that a regression toward the mean will occur and the needs of high achieving students will not be met. As was shown, the results for the state of Kansas do not indicate a regression toward the mean. In fact, as the years have passed more students are meeting the requirements to be categorized as *Advanced* or *Exemplary* and fewer students are being placed in *Unsatisfactory* or *Basic*. The study reported here added to the literature on this as well.

This examination of the data does not provide the rationale for the increase in student achievement in Kansas. While it appears that NCLB has had a positive impact, these data alone don't tell the full story. For example, questions about teaching strategies, local focus on state curriculum, and teacher education could be examined to provide a better understanding of what has actually occurred in Kansas over the past six years to encourage such strong growth in achievement. It is suggested that future research that addresses questions surrounding changes in teaching and learning be studied in relation to increases in student achievement.

Linn (2003) noted that low achieving schools can be disadvantaged by the policies surrounding adequate yearly progress and No Child Left Behind. These schools must not only address progress immediately, they must also strive to meet state directed achievement targets that essentially require low achieving schools to make greater gains than high achieving schools. This could lead this group of schools to implement instructional practices that focus more knowledge level skills and less on complex or higher order thinking tasks in efforts to meet the AYP requirements. Further achievement research could provide information in this area by studying effect size estimates for schools grouped by achievement data for growth across time.

The results reported here are for one state across six years of time in which curricular standards and assessment forms remained constant. Based on these two constants, Kansas is indeed moving toward the No Child Left Behind goal of all students being at proficient or above by 2014. For the 2005-2006 academic year, Kansas adopted new curricular standards and new assessments (KSDE, 2005). As per NCLB requirements, Kansas is now assessing all students in grades 3-8 and once in high school

in both reading and mathematics. Only time will tell if students will continue to show achievement gains on the new curricular standards and state assessments.

References

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Hall, D., Wiener, R., & Carey, K. (2003). *What new "AYP" information tells us about schools, states, and public education.* Washington, DC: The Education Trust.

Kafer, K. (2004). *No Child Left Behind: Where do we go from here?* Washington, DC: The Heritage Foundation.

Kansas State Board of Education (1999a). *Kansas State Mathematics Curriculum Standards*. Retrieved July 21, 2005 from http://www.ksde.org/outcomes/mathstd1999.html

Kansas State Board of Education (1999b). *Kansas State Reading Curriculum Standards*. Retrieved July 21, 2005 from http://www.ksde.org/outcomes/read98.pdf

Kansas State Board of Education (2004). *Kansas adequate yearly progress revised guidance*. Retrieved July 21, 2005 from http://www.ksde.org/sfp/nclb/ayp/rev_ayp_guid.pdf

Kansas State Department of Education (2005). *Kansas Assessments*. Retrieved July 2005 from http://www.ksde.org/assessment/index.html

Linn, R.L., Baker, E.L., & Betebenner, D.W. (2002). Accountability systems: Implications of requirements of the No Child Left Behind Act of 2001. *Educational Researcher 31*(6), 3-16.

Linn, R.L. & Huag, C. (2002). Stability of school-building accountability scores and gains. *Educational Evaluation and Policy Analysis 24*(1), 29-36

Linn, R.L. (2003, Winter). Requirements for measuring adequate yearly progress. *CRESST Issue Brief, 6*. Center for Research on Evaluation, Standards, and Student Testing Los Angeles, CA: Author.

Linn, R.L. (2005) Conflicting demands of No Child Left Behind and state systems: Mixed messages about school performance. *Education Policy Analysis Archives 13*(33). Retrieved July 21, 2005 from http://epaa.asu.edu/epaa/v13n33.

Linn, R.L. (2005, Summer). Fixing the NCLB accountability system. *CRESST Issue Brief, 8*. Center for Research on Evaluation, Standards, and Student Testing Los Angeles, CA: Author.

No Child Left Behind Act of 2001, Pub. L. No. 107-110, 115 Stat.1425 (2002).

Orlich, D.C. (2003) An examination of the longitudinal effect of the Washington Assessment of Student Learning (WASL) on student achievement. *Education Policy Analysis Archives, 11*(18). Retrieved July 21, 2005 from http://epaa.asu.edu/epaa/v11n18/.

Sanders, W.L. (2003). *Beyond No Child Left Behind*. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.